Session G: Numerical Integration

G.1 Introduction

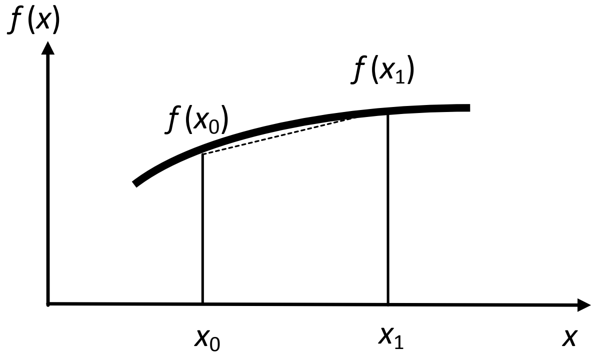
In the last session we looked at the problem of differentiation using finite differences. In this session we turn to what is perhaps a slightly more familiar topic: evaluating integrals between specified limits:

$$I = \int_a^b f(x)dx$$

Generally speaking, the integral of a function means evaluating the area under a curve. Numerical integration is required when the true function, $f(x)$, is either (a) not known, but its value is given at specific points, or (b) is known, but the integral cannot be found analytically. There are a number of different approaches to this problem.

G.2 Trapezoidal Rule

This is by far the simplest scheme of all. Here, the integral of the function between two points, x_0 and x_1, is approximated by the equivalent trapezium formed by drawing a straight line between the points $f(x_0)$ and $f(x_1)$.



The integral is straightforwardly given by:

$$I = \int_{x_0}^{x_1} f(x)dx \approx 0.5[f(x_0) + f(x_1)][x_1 - x_0]$$

If we have data defined at specific points between two limits, $x = a$ and $x = b$, the integral is obtained simply by summing over all the panels:

$$I = \int_a^b f(x)dx \approx \sum_{i=0}^{i=n-1} 0.5[f(x_i) + f(x_{i+1})][x_{i+1} - x_i]$$

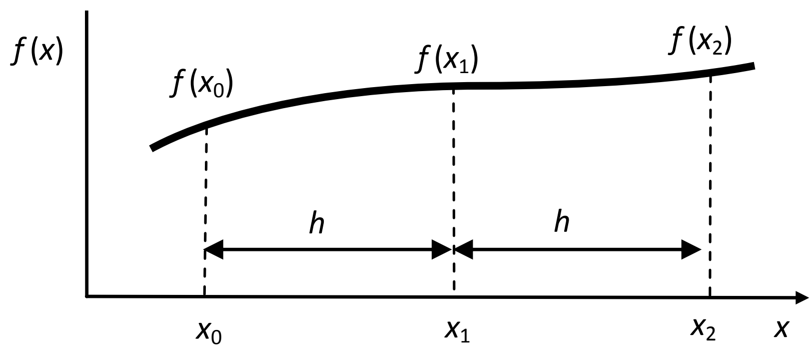
where $x_0 = a$ and $x_n = b$. Note that the panels do not have to be of equal width. If, on the other hand, they are of the same width, h, then the integral is given by:

$$I = \int_a^b f(x)dx \approx \sum_{i=0}^{i=n-1} 0.5[f(x_i) + f(x_{i+1})]h$$

Obviously, the trapezoidal rule is inaccurate to the extent that the actual curve $f(x)$ varies from the straight line drawn between points $f(x_i)$ and $f(x_{i+1})$.

G.3 Simpson's Rule

This method is also simple in concept and, apart from the case where $f(x)$ is a straight line, gives results more accurate than the trapezoidal rule. It provides acceptable accuracy in most general situations. the method is equivalent to simply fitting a parabola through three adjacent points in $f(x)$, say $f(x_0)$, $f(x_1)$ and $f(x_2)$:



Notice that I have assumed the panels are of equal width, h. Using **central** differences to second order, which is good for a parabola, we can write:

$$f(a_0 + mh) = f(a_0) + \delta\{f(a_0)\}m + \delta^2\{f(a_0)\}m^2/2!$$

where

$$\delta\{f(a_0)\} = [f(a_0 + h/2) - f(a_0 - h/2)]$$

and

$$\delta^2\{f(a_0)\} = [f(a_0 + h) - 2f(a_0) + f(a_0 - h)]$$

Do not worry for the moment that we do not have $f(x)$ evaluated at the half way points within panels. The area of the first two panels between x_0 and x_2 is given by:

$$I = \int_{x_0}^{x_2} f(x)dx$$

$$I = \int_{m=-1}^{m=1} f(a_0 + mh)\frac{dx}{dm}dm$$

We have effectively introduced a change of variable. We have based our integration about the central point, so $a_0 = x_1$. Also, because a change in m by 1 gives us a change in x of h, it follows that $dx/dm = h$:

$$I = \int_{m=-1}^{m=1} f(a_0 + mh)hdm$$

Now substituting for $f(a_0 + mh)$ from above and integrating with respect to m:

$$I = h \int_{m=-1}^{m=1} [f(a_0 + \delta\{f(a_0)\}m + \delta^2\{f(a_0)\}m^2/2] dm$$
$$I = h \left[f(a_0)m + \delta\{f(a_0)\}m^2/2 + \delta^2\{f(a_0)\}m^3/6\right]_{m=-1}^{m+=1}$$
$$I = h \left[2f(a_0) + (1/3)\delta^2\{f(a_0)\}\right]$$

Notice that the δ term has cancelled, which is why we do not need to worry about $f(x_{1.5})$ etc.

$$I = h[(1/3)f(x_0) + (4/3)f(x_1) + (1/3)f(x_2)]$$

We can now sum over a succession of panels:

$$I = \frac{2h}{3}\left[\frac{1}{2}f(x_0) + f(x_2) + f(x_4) + \cdots + \frac{1}{2}f(x_{2n})\right] + \frac{4h}{3}\left[f(x_1) + f(x_3) + f(x_5) + \cdots + f(x_{2n-1})\right]$$

or written a different way:

$$I = (h/3)[f(x_0) + 4 \sum_{\text{odd index terms}} + 2 \sum_{\text{even index terms}} + f(x_{2n})]$$

What do we do if we have an odd number of panels? The answer is to use the trapezoidal rule for the first or last panel, choosing the one that has the better approximation to a straight line to be done by the trapezoidal rule.

Exercise G.1

1. Compute $\int x \tan(x) dx$ for $x = 0°$ to $21°$ using the Trapezoidal Rule at intervals of $1°$.
2. Repeat using Simpson's Rule.
3. Compare the accuracy of the two methods.

G.4 Gaussian Quadrature

We shall now deal with the most accurate method of evaluating integrals. In both of the previous methods, we have approximated the integral by the sum of function values at equally spaced points. Karl Friedrich Gauss (1755–1855) suggested that the accuracy of numerical integration could be significantly improved by optimising the position of the function values, $y_i = f(x_i)$. To make the best use of this approach, we require that our function $f(x)$ is known analytically, so that we are able to evaluate y at any value of x. The fact that the function is analytic does not guarantee that its integral is also analytic.

Gaussian quadrature algorithms are usually expressed as integrals over the range -1 to $+1$. (Note that *quadrature* is the technical term for numerical integration.) The method depends on replacing the integral over x by a summation of terms involving the integrand $f(x)$ and a weighting function, ω_k, obtained at quite specific points; *not* at regular intervals as in Simpson's Rule:

$$I = \int_{-1}^{1} f(x)dx \approx \sum_{k=1}^{n} f(x_k)\omega_k$$

Don't worry for now that the integral runs from -1 to $+1$. I'll show you how to deal with this later.

How many sampling points do we really need to accurately describe a particular function? Well, that depends! A polynomial of degree N has $N+1$ coefficients. By allowing ourselves to vary the points at which we evaluate our function we effectively double the number of degrees of freedom at our disposal. Hence with n points, we can uniquely determine $2n$ parameters, or, in other words, we can describe *exactly* a polynomial of order $N = 2n - 1$. Remember that adding more points does not necessarily mean higher accuracy. You don't improve the accuracy if the true function is of lower order than the one we are using to approximate it. The extra terms will simply end up with zero (or very small) coefficients.

G.4.1 Calculating the sampling points and their weights

If the approximation to the continuous function involves n points, the method has $2n$ parameters to be determined: x_k and ω_k, where $k = 1, \dots n$. Since we now have a symmetric range. It turns out that the sampling points (and weights) are the same for $\pm x$. Thus, in our approximation the integral becomes:

$$I = \int_{-1}^{1} f(x)dx \approx f(0)\omega_0 + \sum_{k=1}^{m} \left[f(x_k) + f(-x_k)\right]\omega_k$$

for odd n, with $m = (n-1)/2$, or:

$$I = \int_{-1}^{1} f(x)dx \approx \sum_{k=1}^{m} \left[f(x_k) + f(-x_k)\right]\omega_k$$

for even n, with $m = n/2$.

The question now arises how to choose the sampling points, x_k, and the weights, ω_k, so as to achieve the optimal accuracy for the method. In the case of two sampling points, we have four parameters and therefore expect the integral to be exact for a third order polynomial or less. We can apply this condition successively to the functions $1, x, x^2$ and x^3 to determine the points on the abscissa and their corresponding weights. It is straightforward to show that:

$$\int_{-1}^{1} 1dx = 2 = \omega_1\left[1+1\right]$$
$$\int_{-1}^{1} xdx = 0 = \omega_1\left[x_1 - x_1\right]$$
$$\int_{-1}^{1} x^2dx = \frac{2}{3} = \omega_1\left[x_1^2 + x_1^2\right]$$
$$\int_{-1}^{1} x^3dx = 0 = \omega_1\left[x_1^3 - x_1^3\right]$$

Thus, for the case of two points, the abscissa are $x_1 = \pm 1/\sqrt{3}$ and the weight is $\omega_1 = 1$. This is sufficient to determine the proper integral (that is, when both limits are finite) of any polynomial of order 3 or less. Once you have gone through the example below, you might like to verify that this remarkably simple approximation does indeed give the *exact* result for *any* polynomial of order 3 or less that you choose.

Note that the sampling points (or abscissa) are not rational. Note also that the calculation of the abscissa and their weights is independent of the functions we chose. The functions $1, x, x^2$ and x^3 were the simplest we could have chosen. Try a different set and you will see what I mean. Put another way, the abscissa and their weights for a given number of samples are constant. You can find their values tabulated in books of mathematical functions.

The method we have described is the **Gauss-Legendre** method. Tables of the abscissa and their corresponding weights for $n = 2$ (which we just calculated) to $n = 4$ are given below:

	x_i	ω_i
$n = 2$	± 0.577350269189626	1.000000000000000

	x_i	ω_i
$n = 3$	0.000000000000000	0.888888888888888
	± 0.774596669241483	0.555555555555555

	x_i	ω_i
$n = 4$	± 0.339981043584856	0.652145154862546
	± 0.861136311594053	0.347854845137454

To help illustrate the method, let us evaluate an integral for which we know the result:

$$I = \int_0^{\pi/2} \sin(\theta) d\theta = -\cos(\theta)\Big|_0^{\pi/2} = 1$$

We must first normalise our function so that the integral spans the range -1 to $+1$. This requires a change of variable:

$$\int_a^b f(\theta) d\theta = \int_{-1}^1 f(x) \frac{d\theta}{dx} dx$$

This change of variable is often the trickiest part of the method. In this case it is straightforward to show that:

$$\theta = \frac{\pi}{4}(1 + x)$$

gives the correct limits when $x = -1$ and $x = +1$. From this expression we find:

$$d\theta = \frac{\pi}{4} dx \qquad \text{or} \qquad \frac{d\theta}{dx} = \frac{\pi}{4} \quad \text{a constant}$$

Using the Gauss-Legendre quadrature method with $n = 2$, we can approximate the integral as:

$$\begin{aligned}
I &= \int_0^{\pi/2} \sin(\theta) d\theta \\
&= \frac{d\theta}{dx} \int_{-1}^{1} \sin\left(\frac{\pi}{4}(1+x)\right) dx \\
&= \frac{\pi}{4} \int_{-1}^{1} \sin\left(\frac{\pi}{4}(1+x)\right) dx \\
&\approx \frac{\pi}{4} \sum_{k=1}^{n} \sin\left(\frac{\pi}{4}(1+x_k)\right) \omega_k \\
&\approx \frac{\pi}{4} \left[\sin\left(\frac{\pi}{4}(1+0.57735)\right) \times 1.000 + \sin\left(\frac{\pi}{4}(1-0.57735)\right) \times 1.000\right] \\
&= 0.9985
\end{aligned}$$

The result is close to the desired value of unity. The extent to which the numerical result differs from the analytic solution reflects how well a third order polynomial represents a sine function over this particular range.

Exercise G.2

1. Compute $\int \sin(x) dx$ for $x = 0^\circ$ to 90° using the trapezoidal rule with 2, 3, and 4 panels.
2. Repeat using Gaussian Quadrature for $n = 3$ and 4.
3. Compare the accuracy of the two methods and also compare the results for the Quadrature method for $n = 2$ in the example above with those for $n = 3$ and $n = 4$.

Full tables of (x_i, ω_i) data for these and other quadrature formulae are given in "The Handbook of Mathematical Functions (Dover Press)". Also see below for a Fortran routine to calculate them.

Some other quadrature schemes that you should be aware of, include:

Gauss-Laguerre

$$\int_0^{\infty} \exp(-x) f(x) dx = \sum_{k=1}^{n} f(x_k)\omega_k$$

Gauss-Hermite

$$\int_{-\infty}^{\infty} \exp(-x^2) f(x) dx = \sum_{k=1}^{n} f(x_k)\omega_k$$

These allow you to evaluate improper integrals, where one or both of the limits are infinite. Each scheme has a different set of weights and sampling points. The names associated with each method are associated with a specific type of polynomial. It turns out that the sampling points are simply the zeros of the corresponding polynomials. In the case of the $n = 2$ Gauss-Legendre method for example, $x = \pm 1/\sqrt{3}$ are the roots of the second order Legendre polynomial.

Appendix

The book *Numerical Recipes*, (Fortran Version) by W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, Cambridge University Press 1989, has a subroutine listed which calculates the Gaussian quadrature x_i and ω_i values (called the abscissa and weights in the program). Note that it will provide these for any starting and ending limits to the integration, not just -1 to $+1$. However, you can check the values in the tables above using the routine and by providing -1 and $+1$ as the limits to the integration. The values should agree with the tables above to within the last few digits.